

# **Быстрая классификация JPEG – изображений**

Дорогов А.Ю., Курбанов Р.Г., Разин В.В.

Санкт-Петербургский государственный  
электротехнический университет (СПбГЭТУ) "ЛЭТИ"  
dorogov@lens.spb.ru

## **Аннотация**

Среди различных видов классификаций мультимедиа изображений наиболее значимой и наиболее сложной является задача их семантической классификации [1,2]. По статистическим оценкам в мультимедийных базах данных большинство изображений (до 80%) представлены в JPEG формате или производных от него (JFIF, SPIFF, JBIG, JPEG-EXIF, MPEG). Формат сжатия JPEG обеспечивает значительную экономию ресурсов, как при хранении, так и при обработке зрительных образов, что и определяет его широкую распространенность. Поэтому разработка эффективных методов семантической классификации JPEG-изображений является в настоящее время весьма актуальной задачей.

В предлагаемом подходе используются три основные идеи: 1) для классификации изображений используется спектральное признаковое пространство, формируемое стандартной процедурой блочного кодирования JPEG-формата, что позволяет производить классификацию без восстановления изображения. 2) Семантика полного изображения является производной от семантики сегментов изображения, что позволяет реализовать экономную иерархическую процедуру классификации. 3) Исключаются какие-либо априорные предположения о конфигурации семантического класса в пространстве признаков, классификация выполняется по достоверным прецедентам базы данных, что позволяет ограничиться достаточно простыми метрическими методами классификации.

В результате выполнения проекта были разработаны алгоритмы адаптивной сегментации изображений, алгоритмы информативной оценки системы первичных признаков и формирования сложных вторичных признаков, алгоритмы нечеткой метрической

классификации сегментов изображения, алгоритмы нечеткой семантической классификации видеообразов по результатам сегментной классификации. Алгоритмы реализованы в программной среде МАТЛАБ и представляют собой интерактивную программу накопления, обучения и классификации видеоданных представленных в JPEG-формате.

Проведенные тестовые испытания в целом подтверждают эффективность предложенных подходов. Разработанные алгоритмы могут быть использованы для семантической классификации при достаточной представительности базы данных. Надежность распознавания может быть улучшена добавлением уровня онтологий и расширением вторичного признакового пространства.

### **Abstract**

Between different kinds of classifications the most complex task is the semantic classification of video patterns [1,2]. Accordingly to statistic estimations in multimedia data bases the majority of images (until 80 percents) are represented in JPEG format or in derivations from one. JPEG compressing format provides the significant resource economy the both storing and processing of video images and it determinates its widely propagation. Therefore the developing of effective methods of semantic classification for JPEG images is present-day task.

In the project the three main ideas are used. 1) The spectral indication space formed with standard JPEG procedure is used for image classification. It permits to make classification without full image restoration. 2) The semantic of a total image is carried out from image segment semantics that permits to realize the economic hierarchical classification procedure. 3) Any a priory suppositions about point locations of a semantic class are excluded. It permits to use the enough simple metric classification methods.

In the project there were developed the algorithms of adaptive image segmentation, the algorithms of informative estimation of primary indications and the algorithms of forming complex second indications, the algorithms of fuzzy metric classification for image segments, the algorithms fuzzy image semantic classification on base of the image segment classification. All algorithms are realized in program MATHLAB media and ones represent the interactive program for accumulation, teaching and classification of JPEG-images.

The carried out experiments confirm the effectiveness of supposed methods. Developed algorithms may be used for semantic classification

in condition of representative data base. The reliability of image recognition may be increased with adding of ontology level and extending the space of second indications.

## **1. Введение и обзор ключевых работ по исследуемой тематике**

**1.1. Постановка задачи.** JPEG формат представляет собой один из лучших методов сжатия с потерями, в котором обобщен полувекторный опыт исследований инженеров и ученых, работающих в компьютерной, телевизионной и других областях, связанных с человеческим зрением и компьютерной графикой. В мультимедиа и Интернет технологиях этот формат широко используется для хранения, обработки и передачи изображений по каналам связи. Типичной задачей обработки мультимедиа изображений является их классификация. Среди различных видов классификаций по уровню значимости и уровню сложности выделяется задача семантической классификации зрительных образов, позволяющая получить содержательный для человека ответ на вопрос: «Что изображено на картинке?». Прогресс в распознавании семантики образов, оказывает непосредственное воздействие на развитие систем компьютерного зрения, робототехнических систем, поисковых систем Интернет технологий, специализированных баз данных для мультимедиа систем. Время, затраченное на классификацию образов, имеет решающее значение как для баз данных большого объема, так и для систем работающих в реальном масштабе времени. Высокое быстродействие систем классификации может быть достигнуто за счет сокращения непродуктивных преобразований видеобраза и использования быстрых алгоритмов обработки.

Анализ семантики видеобраза является вершиной иерархической процедуры обработки изображений. В основании пирамиды лежат методы формирования первичной системы информативных признаков. Основное требование на данном этапе – обеспечить максимально-возможную инвариантность признаков к топологическим преобразованиям и высокое быстродействие в получении первичной информации. Первичные признаки представлены в пространстве высокой размерности и имеют большие различия по уровню информативности, эти обстоятельства препятствуют их непосредственному использованию в задаче классификации.

На следующем уровне иерархии формируется система вторичных признаков с примерно одинаковым уровнем значимости. Исходной информацией при этом служит анализ накопленной базы данных, который производится всякий раз при добавлении новых данных. Главное требование к процедуре анализа состоит в том, чтобы подобрать оптимальную систему признаков и сократить размерность признакового пространства. Временные затраты на анализ базы данных не имеют решающего значения, поскольку характерны только на этапе обучения системы. В режиме обучения классифицирующая система, взаимодействуя с оператором, накапливает банк семантических понятий.

На верхнем уровне пирамиды, параметрическими или непараметрическими методами решается задача семантической классификации образов. Ответ может быть многозначным, поэтому необходимо ранжировать полученные решения, используя ту или иную оценку уровня значимости. Такова общая схема решения задачи семантической классификации.

**1.2. Обзор существующих методов решения.** Известные методы решения задачи, семантической классификации изображения, сохраняя в целом общую схему, имеют большое разнообразие в способах и стратегиях реализации этапов.

Проблема распознавания семантики изображения (в зарубежной литературе используется аббревиатура CBIR – Content-based image retrieval) имеет более чем 20-летнюю историю. Первые работы относятся к 80-м годам прошлого века [3]. Однако наиболее существенное развитие это направление получило в последние несколько лет. Многие из выполненных исследований посвящены развитию стратегий последовательного уточнения запроса и оптимизаций поисковых процедур для изображений в конкретных базах данных большого объема [4,5,6,7,8]. Хорошо известным продуктом является система анализа семантики MARS, разработанная в университете Illinois [4]. К другим средствам относятся системы анализа PicToSeek [6], DrawSearch [7] и Viper [8]. Общая стратегия в разработке таких систем семантического поиска состоит в создании нового запроса, который оптимизируется в процессе диалога с пользователем. Однако для эффективного использования эта стратегия требует сложной трансформации базы данных в лингвистическую модель с взвешенным набором весов для терминологических переменных. Кроме того, в некоторых системах (Viper [8]) генерации подобной модели приводит к очень большой размерности признакового пространства. Процедуры рафинирования поискового запроса использует следующие методы:

многомерные индексные структуры [9], набор признаков, выделяемых из примеров, представляющих интерес для пользователя [10], и взвешенные средние для позитивных и негативных примеров [11]. Интерактивный диалог с пользователем в CBIR системах трактуется как парадигма супервизорного обучения, стимулирующая механизм человеческого восприятия образов. В качестве средства формирования весов используются супервизорные нейронные сети [12], методы вероятностной классификации [13] и взвешенные Евклидовы расстояния [14] в многомерном пространстве.

В процессе диалога производится назначение весов терминологическим понятиям. В результате строится некоторая функция подобия. Однако разделяющая мощность такой функции сильно ограничена, поскольку из-за проблем вычислительной сложности для построения функций подобия используется, как правило, не более чем квадратичная форма. Нейросетевые модели хорошо подходят для решения данной задачи, но требуют наличия большого объема обучающих данных для каждого нового запроса.

Альтернативный подход несупервизорного обучения, основанный на использовании самоорганизующихся карт Кохонена, рассматривается в работе [15]. Достоинство этого метода в том, что в процессе обучения системы не требуется взаимодействия с человеком-оператором. Но вследствие ограниченности правил формирования карт это с другой стороны приводит к снижению разделяющей мощности системы распознавания. Кроме того, для восприятия результата классификации человеком необходимо выполнить трактовку автоматически вырабатываемых семантических понятий, что не всегда возможно.

Задача семантического распознавания пересекается с задачей создания искусственного разума [16,17]. Это направление развивается в нашей стране в Институт проблем управления РАН им. В.А. Трапезникова. Сущность подхода состоит в трансформации количественной информации любого вида к «текстовой форме» (точнее к номинальной системе признаков). В процессе не супервизорного обучения (по определенным ценовым алгоритмам) реализуется выделение повторяющихся текстовых цепочек, которые трактуются как семантические понятия. Главная проблема использования данного подхода связана с адекватной трансформацией изображения к «текстовому» образу.

### **1.3. Конкретизация целей проводимого исследования.**

Приведенный обзор показывает, что в настоящее время научно-техническое направление семантической классификации

изображений активно развивается, но каждый из известных методов имеет как определенные достоинства, так и недостатки. Наилучшие результаты следует ожидать при решении конкретных задач. В нашей работе предлагается ограничиться задачей семантической классификации только JPEG изображений. JPEG-сжатие основано на использовании априорных данных о свойствах человеческого зрения, поэтому следует ожидать, что кодированный образ будет представлен в признаковом пространстве с достаточно высокой степенью информативности.

В качестве первичной системы признаков предлагается использовать спектральные коэффициенты блочного кодирования. Размерность пространства первичных признаков существенно ограничена за счет используемой в JPEG схемы адаптивного квантования. Таким образом, снимается проблема формирования первичной системы признаков и устраняется необходимость их прореживания по уровню информативности, это обеспечивает принципиальную основу для построения быстрых процедур семантической классификации. Главной целью настоящего исследования является разработка методов и алгоритмов семантической классификации JPEG-изображений.

## **2. Идея исследования**

В JPEG формате цветные изображения представляются в виде яркостной и двух цветоразностных компонент. Алгоритм JPEG-сжатия включает в себя два этапа. На первом этапе выполняются разделение изображения на блоки размеров  $8 \times 8$  пикселей, которые подвергаются двумерному косинусному преобразованию. В результате получаются матрицы спектральных коэффициентов размером  $8 \times 8$ . Было доказано, что косинусное преобразование является оптимальным для случайных процессов марковского типа. Случайные процессы этого класса в статистическом смысле хорошо аппроксимируют случайное множество изображений. Это означает, что, обеспечивая в среднем достаточно хорошее качество восстановленного изображения, можно ограничиться только несколькими спектральными коэффициентами и отбросить ряд высокочастотных. В результате объем передаваемой информации значительно сокращается, причем оставшиеся коэффициенты несут наиболее существенную информацию об изображении. Прореживание коэффициентов реализуется за счет процедуры нелинейного квантования при переходе к 8 (или 12) битному представлению. Кроме того, при кодировании учитывается также,

что человеческий глаз обладает более низким пространственным разрешением для цветовых перепадов по сравнению с разрешением для перепадов яркости. Поэтому цветоразностные компоненты передаются с вдвое меньшим разрешением, что в два раза сокращает объем передаваемой информации для каждой компоненты. На втором этапе сжатие производится без потерь за счет использования кода Хаффмана. При восстановлении изображения после декодирования Хаффмана восстанавливаются 8-ми (12) битные коды переданных спектральных коэффициентов по яркостной и каждой цветоразностной компоненте. Эти коды можно использовать как первичные информативные признаки, которые уже не требуют какого либо цензурирования по уровню информативности. Таким образом, для классификации изображения нет необходимости в его полном восстановлении, поскольку признаковое пространство уже на промежуточном этапе подготовлено процедурой JPEG-кодирования. Исключение этапа полного восстановления изображения обеспечивает значительную экономию времени при решении задачи классификации видеопictures.

При JPEG-кодировании для каждого блока изображения обязательно сохраняется нулевые спектральные коэффициенты, которые несут информацию о среднем уровне яркостной и цветоразностных компонент. В предлагаемом подходе эта тройка коэффициентов используется для блочной пространственной сегментации изображения. Сегментация реализуется процедурой метрической кластеризации, с метрикой Хэмминга, межклассовое расстояние оценивается по принципу «максимально удаленных соседей». Благодаря процедуре кластеризации в зависимости от уровня сложности изображение может быть разделено на 1-7 сегментов. В каждом сегменте в свою очередь выделяются непрерывные компоненты, число которых не ограничивается. Сегменты описываются в 100 – мерном признаковом пространстве. (принципы формирования вторичных признаков представлены в разделе 3).

В процессе обучения в базе данных накапливаются 100-мерные вектора, каждый из которых соответствует одному сегменту изображения. Оператор присваивает сегменту концептуальное понятие (concept), которое может быть снабжено определением (modifier). Тестовые комбинации modifier/concept определяют семантику сегмента. Таким же образом присваивается семантика всему изображению. В базе данных выявляется частотное отношение между семантикой изображения и семантикой

сегментов. Вторичные признаки в базе данных с помощью энтропийной меры оцениваются по уровню информативности. Полученные данные используются для агрегации вторичных признаков с целью повышения уровня их информативности. В результате на уровне сегментов образуются 20-25 сложных признаков (конкретное значение заведомо не определено и зависит от накопленной базы данных). Все сложные признаки имеют примерно одинаковый уровень информативности. Значение уровня информативности определяет степень доверия к сложному признаку и используется впоследствии для вычисления функций принадлежности семантическому образу.

На уровне образа изображения вне зависимости от сегментного деления формируются еще 18 признаков, которые образуют параллельную систему признаков. Эти признаки также адаптивно агрегируются в сложные признаки. В результате образуется 5-7 сложных признаков уровня образа.

Семантика образа распознается по двум системам признаков. На уровне образов это прямое распознавание, которое производится по каждому сложному признаку уровня образа на основе методов метрической классификации. Результаты накапливаются, и их частотность определяется степень доверия к результатам распознавания.

На уровне сегментов распознавание семантики образа выполняется в два этапа. На первом этапе распознается семантика сегментов изображения подобно тому, как это делается для признаков уровня образа. А на втором этапе на основе найденного ранее по базе данных отношения между семантиками сегментов и образов, определяется функции принадлежности для семантики образов. Данный принцип распознавания отличается от известных иерархическим комплексированием параметрических и непараметрических методов классификации образов.

Результаты по обеим системам признаков объединяются, образуя окончательный результат семантического распознавания. В целом система распознавания неявно образует нечеткую нейронную сеть, в которой выходными лингвистическими переменными являются семантики образов.

Новизна предлагаемых решений заключается в:

- 1) способе получения первичной системы признаков;
- 2) способе адаптивной сегментации изображения;
- 3) способе формирования вторичной системы признаков;
- 4) способе адаптивного агрегирования вторичных информативных признаков;



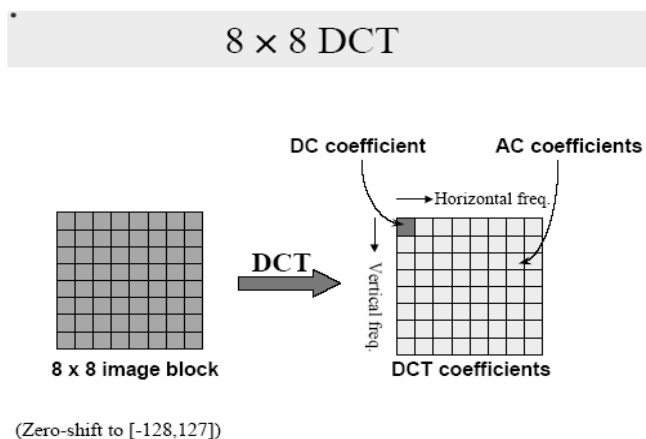
- 5) способе организации базы данных;
- 6) способе иерархической классификации семантических образов.

### 3. Описание методов, алгоритмов и экспериментов

#### 3.1. Формирование первичных информативных признаков

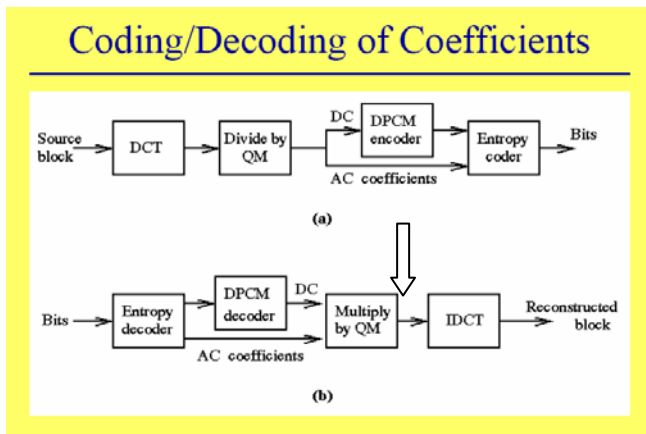
Алгоритм декодирования JPEG-изображения реализован в соответствии с процедурами установленными рекомендациями Т.81, Т.83, Т.84 [18, 19, 20] международного агентства в области коммуникаций ITU-T. Из допустимого множества профилей в рамках проведенного исследования в полном объеме реализован только наиболее распространенный профиль Baseline, поддерживаемый рекомендацией JFIF [21].

Процедура сжатия данных основана на блочном кодировании изображений. Схема блочного кодирования показана на Рис. 1. Общая схема кодирования/декодирования JPEG изображений



**Рис. 1.** Блочное кодирование. DCT – дискретное косинусное преобразование.

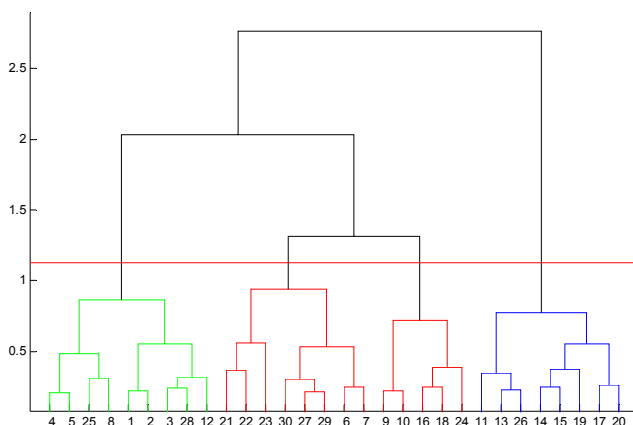
представлена на Рис. 2. Вертикальная стрелка указывает позицию отбора первичных информативных признаков.



**Рис. 2.** DCT, IDCT – прямое и инверсное косинусное преобразование, DC, AC – нулевой спектральный коэффициент и множество высокочастотных коэффициентов, Entropy coder/decoder – кодирование и декодирование Хаффмана, QM – процедуры квантования, деквантования спектральных коэффициентов.

### 3.2. Алгоритм сегментации изображения

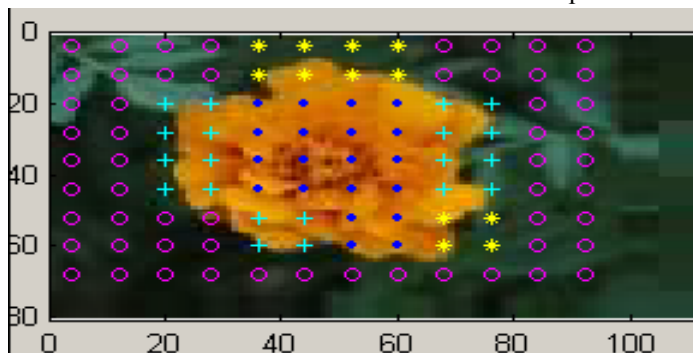
DC-коэффициенты по яркостной и двум цветоразностным компонентам образуют множество векторов размерности 3. Множество кластеризуется метрическим методом, расстояние между векторами определяется выражением:



**Рис. 3.** Дендрограмма сегментации изображения. Горизонтальная линия соответствует адаптивному уровню разделения.

$$L = \sum_{i=1}^3 |x_i - y_i| / D_i,$$

где  $x_i, y_i$  - координаты векторов,  $D_i$  - диапазон вариаций значений по координате. Расстояние между классами оценивается по принципу «максимально-удаленных соседей». Разделение на классы производится адаптивно, линия кластерного раздела на дендрограмме кластерных расстояний проводится, когда отношение двух смежных уровней дендрограммы падает ниже 85%. Это порог обеспечивает результаты, близкие к субъективным решениям человека. На Рис. 3 показана схема образования 4-х



**Рис. 4.** Сегментированное изображение

сегментов изображения. На Рис. 4 показан результат сегментации изображения. Блоки 8\*8 помечены точечными маркерами различного вида, каждый из которых соответствует определенному сегменту изображения.

### 3.3. Формирование вторичных признаков

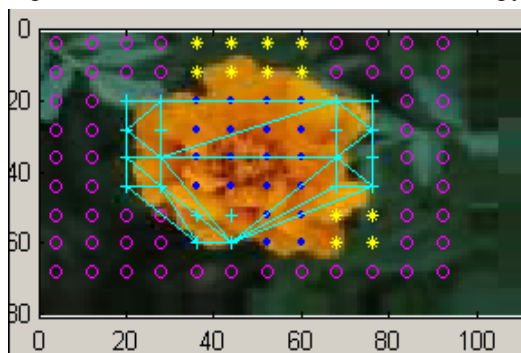
По результатам сегментации формируются вторичные признаки, к которым относятся относительные площади сегментов, логические признаки, характеризующие размещение сегмента в поле изображения, число непрерывных компонент и др. Признаки формы сегмента представлены в Табл. 1. Кроме того, в каждом сегменте выделяются 14 точек (блоков 8\*8) с наибольшими значениями контраста по яркостным и цветоразностным компонентам, пороговый уровень устанавливается адаптивно и его значение также является вторичным признаком. Уровень контраста оценивается энергией высокочастотных компонент косинусного спектра. Выпуклая оболочка контрастных точек (далее называемая «полигон») несет количественную информацию, тесно связанную с семантикой изображения. На Рис. 5 показаны точки полигона для

яркостной компоненты сегмента граничных областей. Точки условно соединены ребрами графа.

**Табл. 1. Информативные признаки сегмента**

Param.potential=s/(row*col);	потенциал сегмента (равен относительной площади, занимаемой сегментом в поле изображения).
Param.YCbCr_DC=CL(k,:);	массив из трех элементов; содержит яркостную и цветоразностные компоненты Y,Cb,Cr для доминирующего фона сегмента.
Param.YCbCr_AC=aYCbCr;	массив из трех элементов: <u>содержит</u> энергетические значения переменных, составляющих яркостной и цветоразностных компонент сегмента.
Param.LPosition=LPosition;	структура, содержащая логические позиции сегмента в поле изображения.
LPosition.top=1	- сегмент касается верхней границы изображения
LPosition.bottom=1	- сегмент касается нижней границы изображения
LPosition.left=1	- сегмент касается левой границы изображения
LPosition.right=1	- сегмент касается правой границы изображения
LPosition.center=1	- сегмент имеет блоки, принадлежащие области центра
LPosition.quadrant1=1	- сегмент имеет блоки, принадлежащие первому квадранту
LPosition.quadrant2=1	- сегмент имеет блоки, принадлежащие второму квадранту
LPosition.quadrant3=1	- сегмент имеет блоки, принадлежащие третьему квадранту
LPosition.quadrant4=1	- сегмент имеет блоки, принадлежащие четвертому квадранту
Param.Narea	- число изолированных областей в сегменте

По полигону строится 12 вторичных признаков, характеризующих его форму. Набор признаков, описывающих полигон представлен в Табл. 2. Каждое поле структуры Polygon



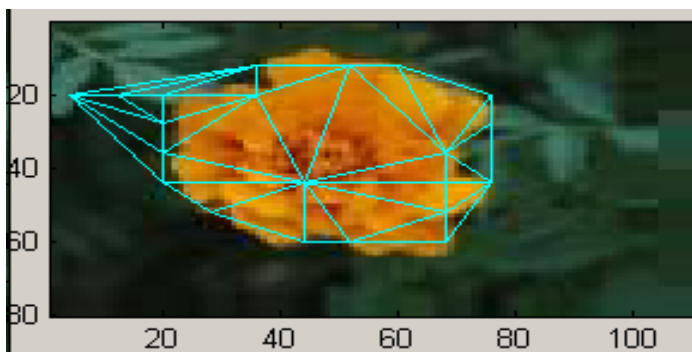
**Рис. 5. Полигон сегмента граничных областей.**

имеет три компоненты. Координаты центра масс и центра полигона представляют собой матрицы размером 3\*2. Все координаты представлены в относительных единицах, нормировка выполняется по отношению к фактическим размерам изображения. Линейные размеры нормируются по отношению к максимальной диагонали изображения.

**Табл. 2. Информативные признаки полигона**

Polygon.Potential=S/(X*Y); % матрица потенциалов полигонов
Polygon.Diameter=D/sqrt(X^2+Y^2); % матрица диаметров полигонов (относительно диагонали отображения)
Polygon.FormFactor =S/D^2 – фактор формы
Polygon.DAngle –углы ориентаций максимального диаметра для полигонов
Polygon.EAngle=EAngle; углы ориентаций вектора главной компоненты полигонов
Polygon.Radius=Radius; максимальные относительные радиусы для полигонов
Polygon.EFactor=EFactor; отношение длины минимальной оси аппроксимирующего эллипса полигона к длине максимальной оси
Polygon.Mass_center=Mass_center./[X,Y]; % матрица относительные координаты центра масс всех точек полигонов
Polygon.Polygon_center=Polygon_center./[X,Y]; % матрица относительных координат геометрического центра полигонов
Polygon.Level=Level – уровни выделения точек полигонов
Polygon.MaxValue – максимальные значения для точек выделения
Polygon.Points=Points – массив ячеек, содержащих координаты точек полигона
Polygon.Graph – массив ячеек, содержащий матрицы смежностей графа Делоне для точек полигона
S – площадь полигона
X*Y – размеры изображения в пикселях
D – длина максимальной диагонали изображения

Подобным образом полигоны строятся для каждой непрерывной компоненты сегмента и для всего изображения в целом. Для изображения в целом полигон строится на 21 точке, а для каждой непрерывной компоненты сегмента - на 7 точках. На Рис. 6 показан полигон уровня изображения для яркостной компоненты.

**Рис. 6. Полигон уровня изображения**

В пространство признаков сегмента в данной версии программы включается только максимальная непрерывная компонента, матрица смежностей не используется и зарезервирована для будущих версий. Всего для количественной характеристики сегмента

используется 100 вторичных признаков. По полигону уровня изображений формируется 39 вторичных признаков. Кроме того, формируются специфичные признаки уровня образа

**Табл. 3. Информативные признаки уровня образа**

Pattern.Y=Y;	- размер образа по вертикали.
Pattern.X=X;	- размер образа по горизонтали.
Pattern.XYratio;	- отношение X/Y
Pattern.Fovea;	0/1 – фовеальный фильтр включен / выключен
Pattern.Veber;	0/1 – выборка точек полигона по перепаду / по контрастности
Pattern.Nseg=d;	- число сегментов в изображении.

представленные в Табл. 3.

В данной версии программы эти признаки не используются для классификации, но сохраняются в базе данных для информационных целей и будущих расширений. Процедура формирования вторичных информативных признаков используется как на этапе накопления базы данных, так и на этапе распознавания.

На этапе накопления базы данных, для каждого сегмента и каждого образа оператор вводит концептуальные понятия (concept) и определения к ним (modifier) в виде текстовых слов. Эта пара слов считается семантикой графического образа или его сегмента. Вместе с вектором признаком эта информация сохраняется в базе данных. По базе данных строится нечеткое отношение Transit между семантикой сегмента и семантикой образа. Значения матрицы нечеткого семантического отношения определяется числом совместного проявления в базе данных семантики сегмента с определенной семантикой образа.

### **3.4. Формирование сложных признаков**

**3.4.1. Оценка информативности.** Формирование сложных признаков выполняется на основе анализа информативности признаков в накопленной базе данных. Для анализа информативности признаков используется мера взаимной информации [22] признака с верифицированной семантикой сегмента и или изображения в целом. Предварительно количественные признаки квантуются с точностью 5% от диапазона изменения, а семантика сегментов и образов кодируется числовыми кодами. В итоге поле признаков приводится к номинальной шкале, где каждое значение можно считать буквой. Для количественных признаков число букв в алфавите равно 20. Для семантики число букв определяется числом различных семантических понятий в базе данных. Оценка информативности признаков выполняется на основе информационной матрицы, по одной координате которой

представлены коды признака, а по другой коды семантического образа, пример подобной матрицы показан ниже.

$x/y$	$a$	$b$	$c$	$d$	$e$	$f$	$g$	$h$	
0	$m_{00}$	$m_{01}$	$m_{02}$	$m_{03}$	$m_{04}$	$m_{05}$	$m_{06}$	$m_{07}$	$m_0$
1	$m_{10}$	$m_{11}$	...					...	$m_1$
2	...	...							$m_2$
3	...	...							$m_3$
4	$m_{40}$	$m_{41}$	...					...	$m_4$
	$n_0$	$n_1$	$n_2$	$n_3$	$n_4$	$n_5$	$n_6$	$n_7$	$M$

Данную матрицу можно рассматривать как накопленную по базе данных статистику работы канала передачи информации. Каждый элемент матрицы указывает на число переходов  $i$ -ой буквы алфавита признака  $x$  в  $j$ -ю букву алфавита признака  $y$ . Числа  $m_0 \dots m_4$  - (суммы по строкам) определяют композицию признака  $x$ . Числа  $n_0 \dots n_7$  - (суммы по столбцам) определяют композицию признака  $y$ . Величина  $M = \sum_j n_j = \sum_i m_i$  равна числу переданных

символов. Энтропии признаков определяются выражениями:

$$H_x = -\sum_i \frac{m_i}{M} \log \frac{m_i}{M} \quad - \quad \text{энтропия входа,}$$

$$H_y = -\sum_j \frac{n_j}{M} \log \frac{n_j}{M} \quad - \quad \text{энтропия выхода,}$$

$$H_{x \otimes y} = -\sum_i \sum_j \frac{m_{ij}}{M} \log \frac{m_{ij}}{M} \quad - \quad \text{энтропия канала,}$$

$$H_{x:y} = H_x + H_y - H_{x \otimes y} \quad - \quad \text{взаимная энтропия.}$$

Известно [22], что  $H_{x:y} \leq H_x$  и  $H_{x:y} \leq H_y$ . Взаимная энтропия характеризует степень похожести признаков  $x$  и  $y$ . Удобно использовать нормированную величину  $S_{x:y} = H_{x:y} * 2 / (H_x + H_y)$ , которая изменяется в диапазоне  $[0,1]$ . Если  $S_{x:y} = 1$  признаки  $x$  и  $y$  совпадают (с точностью до замены букв). Если  $S_{x:y} = 0$  - признаки  $x$  и  $y$  сильно различаются. Величина  $S_{xy}$  - когда  $x$  является измеряемым информативным признаком, а  $y$  - семантикой сегмента, рассматривается как степень надежности признака  $x$  и

используется далее для вычисления функции принадлежности для лингвистических переменных. Данный метод позволяет выполнить оценку информативности одиночных признаков по отношению к семантике сегмента или образа и отбросить слабые признаки.

#### 3.4.2. Агрегирование признаков

Взаимная энтропия используется также для оценки связи между измеряемыми информативными признаками. Признаки независимы, если их взаимная энтропия равна нулю. Агрегация независимых информативных признаков в сложный признак приводит к усилению разделяющей способности по отношению к семантике образа [22]. Поэтому на этапе классификации все признаки подвергаются агрегации. При равном уровне информативности признаки эффект от агрегации тем выше, чем выше степень взаимной независимости признаков. Агрегация выполняется как адаптивная кластеризующая процедура, реализующая последовательное объединение наиболее независимых признаков. Мерой расстояния между признаками  $x$  и  $y$  является величина

**Табл. 4. Информативность признаков**

Признак	Одиночный признак	Сложный признак
FormFa1	0.659	0.8719
EFacto1	0.677	
Potent1	0.673	0.9222
Radius3	0.651	
Mass1y	0.637	
Poly1y	0.643	

$1 - S_{xy}$ . Агрегация выполняется как рекурсивная процедура последовательной кластеризации множества признаков, при этом используется тот же механизм адаптации, что и при сегментировании изображения. В Табл. 4 демонстрируется эффект агрегации на примере 5 признаков полигона образа. Числовые значения в таблице равны степени информативности  $S_{xy}$  одиночных и сложных признаков по отношению к семантике образа. Степень информативности сложного признака далее обозначается символом  $\chi_i$ .

Структура сложных признаков определяется при запуске программы, и после каждого изменения базы данных.



### 3.5. Семантическая классификация

Предъявляемый паттерн разделяется на сегменты, и по каждому сегменту определяются значения сложных признаков, которые представляют собой векторы небольшой размерности.

*3.5.1. Нечеткая классификация сегментов.* Предъявляемый сегмент обрабатывается по той же схеме, что и при формировании базы данных. Текущее значение сложного признака сравнивается со значениями одноименного признака всех верифицированных сегментов базы данных (прецедентами). В результате сравнения определяются стандартные евклидовы расстояния между текущими и верифицированными значениями. Уровень достоверности классификации по данному сложному признаку определяется величиной:

$$\xi_i = 1 - \min_k (d_{ki}) / \max_k (d_{ki}),$$

где  $d_{ki}$  - расстояние текущего значения  $i$ -го признака до  $k$ -го прецедента. Уровень достоверности максимален и равен 1, когда в базе данных существует прецедент, совпадающий с текущим значением признака. Функция принадлежности к прецеденту вычисляется по правилу:

$$\eta_k = \frac{1}{p} \sum_{i=1}^p \min(\xi_i, \chi_i),$$

где  $p$  - число сложных признаков на уровне сегментов,  $\chi_i$  - уровень достоверности сложного признака. Функция принадлежности к семантике  $s$  определяется выражением:

$$\mu_s = \max_{\eta_k \in K_s} (\eta_k),$$

где  $K_s$  - множество прецедентов с семантикой  $s$ . Значения функции  $\mu_s$  - определяют степени принадлежности анализируемого сегмента термам лингвистической переменной «семантика сегмента». На заключительном этапе выделяется терм, которому соответствует максимальное значение функции принадлежности. Семантика этого терма рассматривается как окончательный результат классификации сегмента, а соответствующее значение функции принадлежности определяет уровень его достоверности.

*3.5.2. Нечеткая классификация образов по семантике сегментов.* Семантическая классификация образов выполняется на основе матрицы нечеткого отношения Transit между семантикой сегментов и семантикой образов. Просматриваются семантики классифицированных сегментов, и выполняется вычисление

функции принадлежности к семантике образов, по следующему правилу:

$$\mu_{\text{lm}}(i) = \frac{1}{n \cdot m(i)} \sum_{k=1}^n \mu_s(k) \text{Transit}(i, k);$$

где  $n$  - число сегментов в образе,  $\mu_s(k)$  - уровень достоверности семантики  $k$ -го сегмента,  $m(i) = \max_k (\text{Transit}(i, k))$  - нормирующий множитель. В моделирующей программе результаты семантической классификации образов упорядочиваются по убыванию функции принадлежности и в таком виде выводятся на экран.

*3.5.3. Нечеткая классификация образов по параметрам полигона образа.* По каждому сложному признаку полигона образа выполняется метрическая классификация по всем прецедентам базы данных. Текущее значение сложного признака сравнивается со значениями одноименного признака всех прецедентов. В результате сравнения определяются стандартные евклидовы расстояния между текущими и верифицированными значениями. Уровень достоверности классификации по каждому сложному признаку определяется величиной:

$$\xi_i = 1 - \min_k (d_{ki}) / \max_k (d_{ki}),$$

где  $d_{ki}$  - расстояние текущего значения  $i$ -го признака до  $k$ -го прецедента. Достоверность максимальна и равна 1, когда в базе данных существует прецедент, совпадающий с текущим значением признака. Функция принадлежности к прецеденту вычисляется по правилу:

$$\eta_k = \frac{1}{g} \sum_{i=1}^g \min(\xi_i, \chi_i),$$

где  $g$  - число сложных признаков на уровне сегментов,  $\chi_i$  - уровень достоверности сложного признака. Функция принадлежности к семантике образов  $p$  определяется выражением

$$\mu_p = \max_{\eta_k \in K_p} (\eta_k),$$

где  $K_p$  - множество прецедентов с семантикой  $s$ . Значения функции  $\mu_p$  определяют степени принадлежности анализируемого образа термам лингвистической переменной «семантика образа».

В моделирующей программе результаты семантической классификации образов упорядочиваются по убыванию функции принадлежности и в таком виде выводятся на экран.

**3.5.4. Комплексирование результатов.** Результаты семантической классификации, полученные по полученным по двум параллельным системам признаков, конъюнктивно объединяются. Результирующая функция принадлежности определяется правилом:

$$\mu = \min(\mu_{lm}, \mu_p).$$

Если множество диагнозов пусто или максимальное значение функции принадлежности меньше заданного порога, то моделирующая программа выдает сообщения отказа от классификации. Данная ситуация возникает, когда база данных недостаточно представительна.

### 3.6. Результаты экспериментальных исследований

Для проведения экспериментальных исследований в среде МАТЛАБ была разработана программа классификации JPEG-изображений. Программа позволяет также формировать базу данных и выполнять оценки информативных признаков. Рабочее окно программы показано на Рис. 7. При декодировании

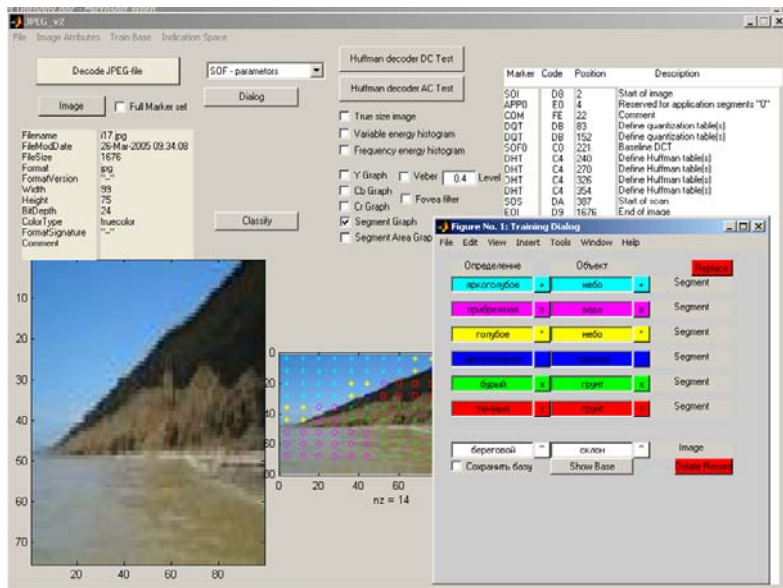


Рис. 7. Скриншот рабочего окна

изображений в правом информационном поле отображаются обнаруженные маркеры JPEG-формата. Основным управляющими элементами интерфейса являются кнопки **Decode JPEG-file** - запуск процедуры декодирования файла, **Dialog** – вызов окна управления базой данных, **Classify** – запуск процедуры классификации. Диалоговое окно, представленное в нижнем правом углу скриншота, реализует ввод семантических понятий и управление базой данных.

Для проведения экспериментов с помощью данной программы была создана база данных содержащая информацию по 100 изображениям. Характеристики базы данных представлены в Табл. 5. Из таблицы видно, что представительность базы данных в среднем не высокая; на каждое семантическое понятие Concept/Modifier на уровне образов приходится 1.5 примера, а для

**Табл. 5. Характеристика базы данных**

Число примеров в базе данных=100
Число сегментов в базе данных=535
Число концептуальных понятий в базе данных=86
Число модификаторов понятий в базе данных=154
Число сочетаний Concept/Modifier на уровне сегментов=214
Число сочетаний Concept/Modifier на уровне образов=67
Средний размер изображений 120*120 пикселей

уровня сегментов 2.5 примера. Однако некоторые концепты представлены достаточно полно. В Табл. 6 показана частотность наиболее представительных концептов на уровне сегментов.

Результаты экспериментов показали, что изображения, которые содержат наиболее представительные концепты базы данных, достаточно уверенно распознаются по семантике.

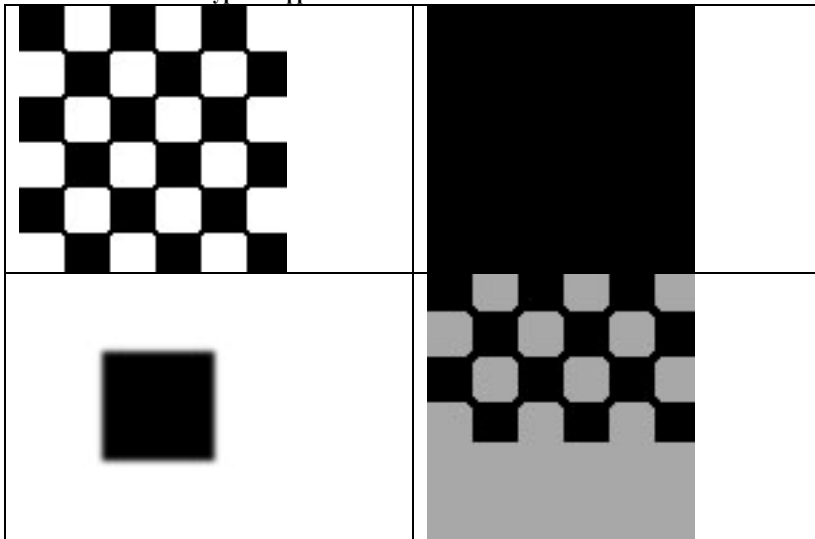
**Табл. 6. Представительность базы данных**

Частотность Concept/Modifier на уровне сегментов	Частотность Concept на уровне сегментов
цветотеневая граница 54	небо 111
текстурный фрагмент 20	граница 81
голубое небо 20	листва 43
белесое небо 17	вода 24
облачное небо 16	грунт 23
светлоголубое небо 16	берег 21
зеленая листва 12	фрагмент 20
темнозеленая листва 12	пятно 17

### Примеры:

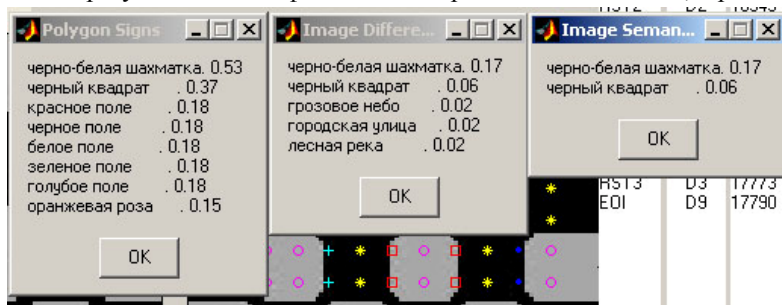
1. *Текстурный фрагмент.* В 1,2,3 квадранте Табл. 7 приведены образы, входящие в базу данных, в четвертом квадранте (правый

**Табл. 7. Текстурный фрагмент**



нижний угол) представлен тестовый образ.

На Рис. 8 показаны скриншот результата классификации. Левое окно – результат классификации по признакам полигона, среднее окно – результат классификации по признакам сегментов, правое



**Рис. 8. Результаты классификации**

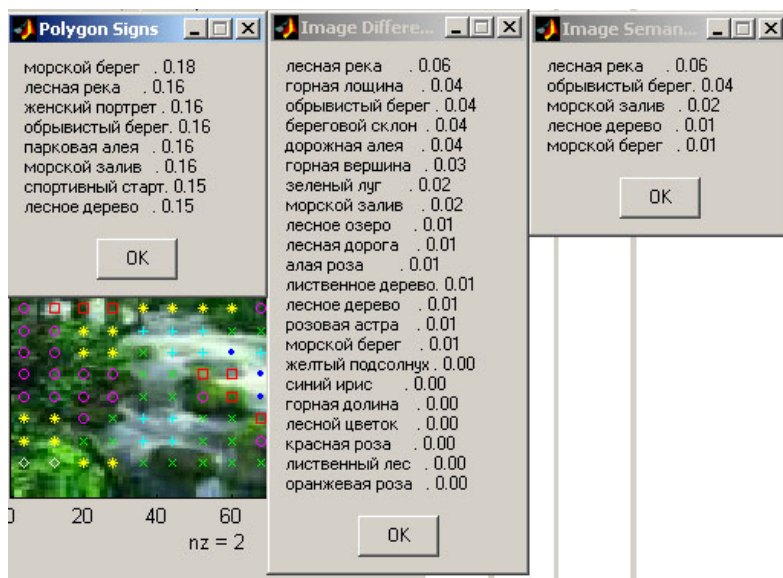
окно – комплексный результат. Числа определяют уровень достоверности классификации.

2. *Лесная река*. В 1,2,3 квадрантах Табл. 8 выборочно показаны примеры, входящие в базу данных, в четвертом квадранте представлено тестовое изображение. Результаты семантической

**Табл. 8. Лесная река**



классификации тестового изображения показаны на Рис. 9. Правое окно соответствует комплексному результату.



**Рис. 9. Результаты классификации образа «Лесная река»**

#### 4. Выводы и обсуждение результатов

Проведенные исследования доказали принципиальную возможность реализации процедуры семантической классификации изображений в признаковом пространстве JPEG-формата. Исключение процедуры полного восстановления изображения является принципиальным решением, которое дает возможность в коммерческих продуктах обеспечить высокое быстродействие классифицирующих подсистем. Для экспериментальных исследований в проекте использовалась программная среда MATLAB, что было обусловлено богатством ее функциональных возможностей в программном и пользовательском интерфейсе и высоким уровнем оперативности при отладке программ. Однако данная среда является интерпретирующей, поэтому скорость выполнения программ существенно ниже, чем в любой компилирующей среде.

Примерные оценки быстродействия дают следующие цифры. Программная процедура, написанная на языке MATLAB, выполняла полное декодирование и сегментацию JPEG-изображения размером 100\*150 пикселей за 10 сек (на процессоре Intel 2,4 ГГц). Семантическая классификация изображения с отображением результата выполнялась за 7 сек. Встроенная функция MATLAB, также реализующая полное JPEG-декодирование, выполняется за 0.1 сек. Поэтому можно ожидать, что время выполнения процедуры семантической классификации в коммерческой программе для данного размера изображения будет также на уровне 0.1 сек.

Представительность базы данных должна быть существенно выше, чем в проведенных экспериментах. Удовлетворительная классификация может быть обеспечена при среднем уровне 15-20 образов на семантический класс. Накопление базы данных процесс достаточно длительный, поскольку связан с действиями оператора. На описание одного изображения (в сегментном представлении) оператором тратится 2-4 минуты. Кроме того, семантическое описание подвержено субъективным взглядам оператора. Субъективность накопленных данных можно уменьшить двумя путями: 1) объединением баз данных созданных различными операторами, 2) добавлением уровня онтологий, определяющего отношения между семантиками. Уровень онтологий может быть выстроен в автоматическом режиме по совпадениям концептов или модификаторов. Развитие уровня онтологии является более перспективным, поскольку позволяет без расширения базы данных

добавить еще один уровень в иерархию семантической классификации, увеличив тем самым достоверность ее работы.

В реализованных алгоритмах размерность первичного признакового пространства была фиксированной и включала в себя признаки образа, сегментов и их полигонов. Однако в базе данных сохраняется также информация о непрерывных областях сегментов, из этой информации в настоящей программе используется только признаки, порождаемые максимальной непрерывной областью. Методы классификации с переменным числом признаков в настоящее время развиты слабо. По оценкам специалистов, их разработка и использование может дать существенный прогресс в решении задачи семантической классификации. Это направление может быть поддержано также развитием методов классификации плоских графов. Для базы данных разработанной программы реализовано сохранение плоских графов, описывающих полигоны (см. рис.5,6). Методы распознавания образов, представленных в виде графов, позволяет обеспечить инвариантность процедур классификации к широкому классу аффинных и топологических преобразований.

К настоящему времени по результатам исследований опубликованы три научные работы [23,24,25] и готовятся статьи для журнала «Искусственный интеллект».

## 5. Литература

1. Smith J.R., S.F.Chang Visualseek: A fully automated content based image query system. In Proc. ACM Multimedia, Boston, MA, Nov,1996.
2. Milind R.N., T.S.Huang Extracting Semantics from Audiovisual Content: The Final Frontier in Multimedia Retrieval.
3. Salton G., McGill M.J. Introduction to Modern Information Retrieval. New York: McGraw-Hill, 1983.
4. Rui Y., Hyang T.S. and Mehrota S. Content-based image retrieval with relevance feedback in MARS, in Proc.IEEE Int. Conf. Image Processing, Santa Barbara, CA, 1997, pp. 815-818.
5. Celentano A., Sciasico E.D. Feature integration and relevance feedback analysis in image similarity evalution. J.Electron. Imaging, vol. 7, no.2, pp. 308-317, 1998.
6. Gevers T., Smeulders A.W.M. PieToSeek: Combining color and shape invariant features for image retrieval. IEEE Trans. Image Processing, vol. 9, pp. 102-119, 2000.



7. Sciascio E.Di., Mongiello. DrawSearch: A tool for interactive content-based image retrieval over the net. Proc. SPIE, vol. 3656, pp. 561-572, 1999.

8. Müller H., Müller S., Marcand-Maillet, and Squire D. McG. Strategies for positive and negative relevance feedback in image retrieval. In Proc. Int. Conf. Pattern Recognition, Barcelona, Spain, 2000.

9. Porkaew S., Mehrotra S., Ortega M. and Chakrabarti K. Similarity search using multiple examples in MARS. In Visual Information and Information Systems. New York: Springer-Verlag, 1999, pp. 68-75.

10. Ciocca G., Schettini R. Using a relevance feedback mechanism to improve content-based image retrieval. In Visual Information and Information Systems, . New York: Springer-Verlag, 1999, pp. 105-114.

11. Muneesawwang P., Guan L. Anonliner RBF model for interactive content-based image retrieval. In Proc. 1<sup>st</sup> IEEE Pasific-Rim Conf. Multimedia, Syney, Australia, Dec. 2000, pp. 188-191.

12. Ei-Nada, Wernick M.N., Yang. Y., Galatsonos N.P. Image retrieval based on similarity learning. In Proc. IEEE. Conf. Image Processing, vol. 3. Vancouver, BC, Canada ,2000, pp.772-775.

13. Peng J., Bhanu B., Oing S. Probabilistic feature relevance learning for content-based image retrieval. Comput. Vision Image Understanding, vol. 75, no. ½ pp. 150-164, 1999.

14. Sclaroff S., Taycher L., Cascia M.L. ImageRover: A content-based image browser for the world wide web. In Proc. IEEE Workshop Content-Based Access Image Video Libraries, Puerto Rico, June 1997, pp. 2-9.

15. Muneesawwang P., Guan L. Automatic Machine Interactions For Content-based Image Retrieval Using a Self-Organizing Tree Map Architecture. IEEE Trans. on Neural Networks, vol. 13, no 4, july 2002.

16. Бодякин В.И. "Куда идешь, Человек? Основы эволюциологии. (информационный подход), М., СИНТЕГ, 1998, С - 332с.

17. Бодякин В.И. "Исследование структурных моделей открытых динамических систем", специальность: 05.13.01 (Управление в технических системах), автореферат диссертации и диссертация на соискание ученой степени к.ф.-м.н., Москва – 1999г.

18. T.81 - Information technology – Digital compression and coding of continuous-tone still images – Requirements and guidelines

19. T.83 - Information technology – Digital compression and coding of continuous-tone still images: Compliance testing.

20. T.84 - Information technology – Digital compression and coding of continuous-tone still images: Extensions.

---

21. JFIF 1.02 (JPEG File Interchange Format).

22. Гоппа В.Д. Введение в алгебраическую теорию информации.– М.: Наука. Физматлит, 1995. –112с.

23. Дорогов А.Ю, Курбанов Р.Г. Шестопапов М.Ю. Нечеткая кластеризация многомерных данных в выборках большого объема // Сборник докладов международной конференции по мягким вычислениям и измерениям SCM'2005, Санкт-Петербург, 27-29 июня 2005.- Т.1.-С.122-126.

24. Дорогов А.Ю, Курбанов Р.Г. Шестопапов М.Ю. Редукция выборки в задачах многомерной кластеризации // Сборник докладов Всероссийской научной конференции Управление и информационные технологии УИТ-2005, Санкт-Петербург 30 июня-2 июля 2005.- Т.1.- С.284-288.

25. Дорогов А.Ю, Краснова А.И., Яшин А.И. Методы управления инвариантностью в задачах распознавания векторизованных изображений // Сборник докладов Всероссийской научной конференции Управление и информационные технологии УИТ-2005, Санкт-Петербург 30 июня-2 июля 2005.- Т.1.- С.275-284.